

**PCT**WORLD INTELLECTUAL PROPERTY  
International Bureau

BD

INTERNATIONAL APPLICATION PUBLISHED UNDER

WO 9605847A1

<b>(51) International Patent Classification <sup>6</sup>:</b> <b>A61K 35/16, 39/00, 39/395</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 96/05847</b> <b>(43) International Publication Date:</b> 29 February 1996 (29.02.96)
<b>(21) International Application Number:</b> PCT/US95/10668 <b>(22) International Filing Date:</b> 22 August 1995 (22.08.95)  <b>(30) Priority Data:</b> 08/294,133 22 August 1994 (22.08.94) US  <b>(60) Parent Application or Grant</b> <b>(63) Related by Continuation</b> US 08/294,133 (CIP) Filed on 22 August 1994 (22.08.94)  <b>(71) Applicant (for all designated States except US):</b> THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA [US/US]; Center for Technology Transfer, Suite 300, 3700 Market Street, Philadelphia, PA 19104-3147 (US).  <b>(72) Inventor; and</b> <b>(75) Inventor/Applicant (for US only):</b> EBERWINE, James [US/US]; Apartment 202, 1025 Clinton Street, Philadelphia, PA 19107 (US).  <b>(74) Agent:</b> LICATA, Jane, Massey; Law Offices of Jane Massey Licata, Woodland Falls Corporate Park, Suite 201, 210 Lake Drive East, Cherry Hill, NJ 08002 (US).		<b>(81) Designated States:</b> CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> A METHOD OF SEQUENCING PROTEINS BY EPITOPE ORDERING AND PROTEIN RESTRICTION MAPPING		
<b>(57) Abstract</b>  Methods of identifying an amino acid sequence of a protein and characterizing post-translational events occurring on the protein through epitope ordering followed by restriction mapping are provided. Modified antibodies containing restriction enzyme sites which are capable of recognizing and binding to amino acids as an epitope or post-translationally modified epitopes are also provided.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LJ	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

**A METHOD OF SEQUENCING PROTEINS BY  
EPITOPE ORDERING AND PROTEIN RESTRICTION MAPPING**

**Background of the Invention**

Automated protein sequencing instruments and  
5 methodologies have advanced considerably in recent years.  
Improvements in this area of technology have resulted in  
increases in the sensitivity, speed and ease in which protein  
sequencing can be performed. Recent advances include the  
development of automated gas-phase-sequencing instruments and  
10 on-line HPLC systems equipped with microbore columns for PTH  
amino acid detection. With the invention of gas-phase  
sequenators, sequence data can be obtained with as little as 10  
to 100 pmol of protein or peptide.

Identification of the N-terminus of proteins and  
15 peptides is important for isolating recombinant DNA clones and  
for characterizing structural and functional protein domains.  
The N-terminal sequences of proteins and peptides are most  
commonly determined using repeated cycles, either manual or  
automated, of the Edman degradation reaction. Each degradation  
20 cycle consists of three steps: coupling, cleavage and  
conversion. In the coupling step, the unmodified N-terminus of  
a peptide or protein is modified with phenylisothiocyanate  
under basic conditions to generate a phenylthiocarbonyl  
peptide. In the second step, the PTC-N-terminal residue is  
25 cleaved from the polypeptide by either liquid or gaseous  
trifluoroacetic acid to form an anilinothiazoline-amino acid  
derivative of the original N-terminal residue and a protein or  
peptide whose penultimate residue now lies at the N-terminus.  
The ATZ-amino acid is unstable and is converted in the third

- 2 -

step by acid to a more stable phenylthiohydantoin (PTH)-amino acid. The N-terminal residue of the (n-1) polypeptide is now available for another cycle of the coupling, conversion, and cleavage steps.

5 In the automated procedure, protein samples are normally adsorbed or coupled to a membrane or glass fiber support that lies in a reaction cartridge. The ATZ-amino acid is washed from the reaction cartridge into a conversion flask where it is converted into the PTH-amino acid. The PTH-amino  
10 acid is then injected into a HPLC column, and the amino acid is detected by UV absorbance and is identified by its retention time. The sensitivity of such analyzers can be as low as 1 pmol of PTH-amino acid. The number of residues which can be sequentially identified is dependent upon the sample amount and  
15 on the sequence itself.

There are several requirements for obtaining sequences from unknown samples using the Edman degradation procedure. First, the sample must be at least 80% pure. Mixtures of proteins generate multiple PTH-amino acids during each cycle of  
20 the Edman degradation. Thus, if multiple proteins are present in similar quantities in a sample, it is almost impossible to assign a sequence to a particular protein. In addition, if the N-terminus is blocked, then sequences can arise from contaminating proteins. Second, the sample must be free of  
25 contaminants such as Tris, glycine, sodium dodecyl sulfate (SDS), or acrylamide which affect the performance of the instrument and create large artifactual peaks in the chromatograms. Finally, there must be sufficient quantity of sample available for analysis. Factors such as sample loss and  
30 N-terminal blockage during purification result in uncertainty in the amount of protein in the sample which can actually be sequenced. Therefore, 10 to 100 pmol of material is normally required for N-terminal sequence analysis via Edman degradation. If the N-terminal sequence is blocked, however,  
35 Edman degradation can not be used to determine the sequence of the protein or peptide.

- 3 -

Monoclonal antibodies have also been used to identify and map specific regions of proteins which act as the epitope.

Kosik et al. *Neuron* 1988, 1, 817-825, disclosed a method of identifying spatially distinct epitopes in the tau molecule by using monoclonal antibodies against the tau molecule. The mean size of the epitope regions identified by the antibodies to the tau molecule were 29.4 amino acids. Cesbron Delauw et al., *Molecular Immunology* 1992, 29(11), 1375-1382 characterized in detail the epitope of the secreted antigen GP28.5 recognized by a mouse monoclonal antibody.

A procedure has now been developed for the easy, sensitive and quick determination of a protein sequence through epitope ordering with antibodies specific for single amino acids, di-amino acids and tri-amino acids followed by restriction mapping. This procedure permits enough protein sequencing to be performed to enable generation of oligonucleotide probes for screening of libraries and to perform PCR to isolate the corresponding cDNA clone. In addition, the epitope ordering procedure allows for characterization of post-translational modification events which occur on a protein.

#### Summary of the Invention

An object of the invention is to provide a method of identifying an amino acid sequence of a protein. This method comprises obtaining antibodies to amino acids. The antibodies are then modified by binding a first nucleotide sequence and a second nucleotide sequence to the antibody, the first nucleotide sequence comprising a modified 5'-nucleotide followed by a single restriction enzyme site and a long nucleotide sequence containing a base pair sequence at the 3' end capable of annealing to a complementary sequence, the second nucleotide sequence comprising the same modified 5'-nucleotide, single restriction enzyme site and long nucleotide sequence as the first nucleotide sequence but having the complementary base pair sequence at the 3' end. The modified antibodies are bound to a protein so that complementary base

- 4 -

pair sequences at the 3' end of said first and second nucleotide sequences of the antibody form a primer template complex which is extended to form a stable double-stranded duplex. The double-stranded duplex comprising bound modified  
5 antibodies is treated with at least one restriction enzyme to digest the duplex and the digested products are separated. A band pattern produced from these digested products is detected and a portion of the amino acid sequence of the protein is identified.

10 Another object of the present invention is to provide a method of characterizing post-translational events on a protein. This method comprises obtaining antibodies to at least two epitopes on a protein with at least one antibody being directed toward a post-translationally modified epitope.  
15 These antibodies are then modified by binding a first nucleotide sequence and a second nucleotide sequence to said antibody, said first nucleotide sequence comprising a modified 5'-nucleotide followed by a single restriction enzyme site and a long nucleotide sequence containing a base pair sequence at  
20 the 3' end capable of annealing to a complementary sequence, said second nucleotide sequence comprising the same modified 5'-nucleotide, single restriction enzyme site and long nucleotide sequence as said first nucleotide sequence and having the complementary base pair sequence at the 3' end. The  
25 modified antibodies are then bound to a protein so that complementary base pair sequences at the 3' end of said first and second nucleotide sequences of said modified antibodies form a primer template complex which is extended to form a stable double-stranded duplex. The double-stranded duplex  
30 comprising bound modified antibodies is treated with at least one restriction enzyme to digest the duplex and the products of the digestion are separated. The band pattern produced from these digested products is detected and a post-translation event on the protein is identified and characterized.

35 Another object of the present invention is to provide modified antibodies which comprise an antibody to a single, bi- or tri-amino acid or an epitope on a protein related to a post-

- 5 -

translational event, a first nucleotide sequence having a modified 5'-nucleotide, a single restriction enzyme site, a long nucleotide sequence, and a first base pair sequence at the 3' end which is capable of annealing to a complementary sequence, and a second nucleotide sequence having the same modified 5'-nucleotide, single restriction enzyme site and long nucleotide sequence as the first nucleotide sequence and a second base pair sequence at the 3' end which is complementary to the first base pair sequence of the first nucleotide sequence, the first and second nucleotide sequences each being bound to the antibody.

#### Brief Description of the Drawings

Figure 1 is an autoradiograph showing hybridization results for use of Ab14 and Abtau1-DNA-spanner complexes in detection of the tau protein. Lane 1 shows a mixture of the Abs which have not been mixed with tau but have been taken through the epitope ordering procedure. Lanes 2 and 3 show the results obtained from two different experiments by mixing tau protein with the Ab-DNA-spanner complexes and going through the epitope ordering procedure with no subsequent restriction enzyme digestion. The BamHI and EcoRI digestions of the material seen in lane 2 are shown in lanes 4 and 5, respectively.

#### Detailed Description of the Invention

A technique has now been developed for the rapid and sensitive identification of amino acid sequences of unknown proteins. This technique can be used to facilitate the generation of nucleic acid hybridization probes. In addition, this technique can be used to characterize post-translational modification events such as amino acid phosphorylation, sulfation, and glycosylation occurring on proteins.

In one embodiment of the present invention, antibodies are prepared against each of the 21 amino acids or to di-amino or tri-amino sequences. Antibodies can also be prepared against non-naturally occurring amino acids such as

- 6 -

selenocysteine, homocysteine and homoserine. In another embodiment, antibodies are prepared against epitopes on a protein relating to post-translational events. For example, antibodies have been prepared against the antigenic di-amino  
5 sequence phospho-tyrosine. In another embodiment, monoclonal antibodies to the tau protein designated as antibody 14 (Ab14) and tau1 (Abtau1) are used. These monoclonal antibodies were raised to two distinct regions of tau with antibody 14 recognizing an epitope in a peptide fragment spanning amino  
10 acids 83-120 of the tau protein, while the Abtau1 antibody is specific for amino acids 131-149. The preparation of these antibodies has been described in Kosik et al. (1988).

As used herein, the term "antibody" is meant to refer to complete, intact antibodies, Fab fragments and F(ab)<sub>2</sub>,  
15 fragments. It is preferred that each specific antibody is made into Fab fragments in accordance with methods well known in the art. The protein structure of complete intact antibodies, Fab fragment and F(ab)<sub>2</sub> fragments and the organization of the genetic sequences that encode such molecules are well known in  
20 the art.

A first nucleotide sequence is prepared which contains a modified 5'-nucleotide followed by a single restriction enzyme site and a long nucleotide sequence which contains a base pair sequence at the 3' end which serves as a sequence to  
25 which a complementary sequence will anneal. A second nucleotide sequence is then prepared comprising the same modified 5'-nucleotide, restriction enzyme site and long nucleotide sequence as the first nucleotide sequence, but having a base sequence at the 3' end in the corresponding  
30 position which is complementary to the base pair sequence of the first nucleotide sequence. In one embodiment, these DNA-spanning oligonucleotides comprises 198 base pairs with four distinct regions. The nucleotide sequence used in accordance with this invention may be conveniently and routinely made  
35 through the well-known technique of solid phase synthesis. Equipment for such synthesis is sold by several vendors including Applied Biosystems (Foster City, CA). Any other

- 7 -

means for such synthesis may also be employed; the actual synthesis of the nucleotide sequences is well within the talents of the routineer. Oligonucleotides are synthesized using standard phosphoamidite chemistry and thus have a 5'-OH group. These oligonucleotides can be phosphorylated using a standard reaction and then derivatized to contain a 5'-NH<sub>2</sub>, using the OligoLink Derivatization Kit (Pierce Chemical). The 5' nucleotide can be modified with the addition of a cystamine containing the free primary amine and is ready to be linked to the antibody. From the 5' end, a nucleotide spacer region is followed by a nucleotide restriction enzyme recognition sequence, which is followed by a second nucleotide spacer region, followed at the 3'-end by a specific nucleotide hybridization sequence. In one embodiment, a 36 nucleotide spacer region is followed by a 6 nucleotide restriction enzyme recognition sequence, followed by a 141 nucleotide spacer region, followed by a specific 15 nucleotide hybridization sequence. In one embodiment, the BamHI and EcoRI restriction sites have been utilized for the DNA-spanners designated BAM36-198, BAM36-198op, RI36-198, and RI38-198op. The BAM and RI spanners have a specific 15 nucleotide sequence at the 3' end which will complement the 15 nucleotide sequence at the 3' end of the 198op sequences.

The nucleotide sequences are attached to a selected antibody covalently using a primary amine cross-linking agent. In one embodiment, the derivatized DNA-spanners were attached to the tau antibodies using glutaraldehyde in accordance with well known methods. However, other bifunctional crosslinking agents such as bis(sulfosuccinimidyl) suberate and dimethyl suberimidate are also useful for forming amide bonds with the primary amine on the 5' end of the nucleotide sequences and the ε-amino group of a lysine residue on the antibody. Such agents are commercially available through companies such as Pierce (Rockford, IL) and their use is well known to those of skill in the art. In one embodiment, Ab14 was complexed with BAM36-198 and BAM36-198op, while Abtau1 was complexed with RI36-198 and RI36-198op using glutaraldehyde. The antibody-oligo conjugates

- 8 -

are subsequently separated from the unincorporated oligonucleotides by protein A column chromatography, in accordance with well known methods. The protein containing fractions are detected by 280 nm absorbance readings and stored  
5 at 4°C.

The modified antibodies are then added to a sample containing a protein of unknown sequence under conditions which discourage annealing of nucleotide sequences. By these conditions it is meant to include, but is not limited to high  
10 salt concentrations or mechanical stabilization of the nucleotide sequence interaction. The modified antibodies then interact with their specific epitopes to bind with them. The ability of Ab14 and Abtaul-DNA-spanner complexes to associate with purified Abtaul protein has been demonstrated for at least  
15 one set of conditions. After binding occurs, T4 DNA polymerase, Klenow and dNTPs are added so that the complementary base sequences of the first and second nucleotide sequences, which are brought into proximity because of the position of the nearest bound modified antibody, form a primer  
20 template complex which is then extended to form a stable double-stranded duplex.

Once the reaction is complete, the chain of connected adjacent modified antibodies is analyzed using restriction enzymes to map the position of the adjacent antibodies. Over  
25 1000 different restriction endonucleases have now been identified with specific cutting sites. Restriction mapping is used frequently to map and characterize DNA fragments by those of skill in the art. Restriction mapping is accomplished with individual enzymes in accordance with the restriction enzyme  
30 suppliers' recommended conditions. For example, EcoRI digestion of the Ab-DNA mixtures is performed by addition of the following reagents in their final concentrations to the mixture: 100 mM Tris pH 7.5, 7 mM MgCl<sub>2</sub>, and 50 mM NaCl. The enzymatic digestion then proceed for two hours at 37°C.  
35 Restriction enzymes can be obtained from a variety of different vendors including Promega (Madison, WI), New England Biolabs (Beverly, MA) and Stratagene (La Jolla, CA). These enzymes

- 9 -

recognize unique palindromic sequences of 4 to 8 base pairs and hydrolyze one phosphodiester bond within the recognition sequence on each DNA strand to yield either a flush or 5' or 3' offset double stranded fragment. They are highly specific for  
5 a given recognition sequence and efficient at cleaving these sequences.

The cleaved fragments are then separated on the basis of their length differences by electrophoresis in agarose or polyacrylamide gels. Gels containing various percentages of  
10 polyacrylamide or agarose (0.5 to 4%) can be used in the separation. The gel is run under conditions which achieve the best separation of fragments. In an electric field, the negatively charged nucleotide sequences migrate toward the anode. This migration is retarded by a matrix which contains  
15 large pores that sieve nucleotide sequences in a fashion so that longer nucleotide sequences migrate more slowly than shorter nucleotide sequences. The sieving properties are proportional to the concentration of agarose or polyacrylamide in the gel with migration being more retarded at higher  
20 concentrations. With polyacrylamide, the sieving is even further controlled by the extent of cross-linking in the polymer. For agarose, the pores are very large and the primary resolution variable is the concentration of the agarose used. Buffers used in these methods are selected in accordance with  
25 their buffering capacity which is needed during electrolysis of the water that occurs during electrophoresis. In one embodiment, undigested, EcoRI or BamHI-digested reactions are mixed with gel loading dye, heated, and electrophoresed on a 4% polyacrylamide gel until the bromphenol blue dye is at the  
30 bottom of the gel.

The Ab-oligo-oligo-Ab complexes are transferred to a nitrocellulose membrane by electroblotting. The nitrocellulose membrane is dried and then prehybridized with a solution containing BSA and salmon sperm DNA to prevent non-specific  
35 binding of the probe to the membrane. The blot is then probed with a radioactively labeled oligonucleotide containing the sequence (TGG)<sub>10</sub> which will hybridize with the DNA in the spacer

- 10 -

region of the DNA-spanner oligonucleotides. This probe is labeled at the 5' end using 5'-polynucleotide kinase and <sup>32</sup>P ATP. After hybridization, the blot is washed to remove unbound probe, and the blot is exposed to X-ray film to detect the size and location of the DNA. Figure 1 shows hybridization results for use of Ab14 and Abtau1-DNA-spanner complexes in detection of the tau protein. These Ab-DNA complexes can be hybridized because of free (unreacted) single stranded DNA-spanners on each antibody. Lane 1 shows a mixture of the Abs which have not been mixed with tau but have been taken through the epitope ordering procedure. Lanes 2 and 3 show the results obtained from two different experiments by mixing tau protein with the Ab-DNA-spanner complexes and going through the epitope ordering procedure with no subsequent restriction enzyme digestion. The BamHI and EcoRI digestions of the material seen in lane 2 are shown in lanes 4 and 5 respectively. The slower mobility of the Ab-DNA complexes seen in lanes 2 and 3 show that the epitope ordering procedure produced a higher molecular weight complex while the restriction enzyme digestions of this complex shown in lanes 4 and 5 show that the restriction enzyme sites which were predicted to be generated were indeed generated.

Alternatively, the antibody can be detected by protein A or a second, labeled antibody in accordance with well known methods.

In general, the banding pattern generated by using a single or multiple restriction enzymes determines the identity of the antibody associated with that restriction site. This information is then used to determine the order of the antibodies and the corresponding sequence of the bound protein. In using the present invention to identify the sequence of proteins, it is preferred that five or more modified antibodies be aligned with an amino acid sequence of a protein to make an oligonucleotide which is then used to clone the molecule from a cDNA or genomic library. The cloned sequence is then used to produce the full-length protein sequence. Such techniques are well-known to those of skill in the art as evidenced by wide use of the Edman degradation where identification of only a few

- 11 -

amino acids is required to be able to generate a probe for isolation of the corresponding clone.

If the present method is being used to characterize post-translational events on the protein, less bound antibodies may be required. In this embodiment, antibodies are prepared against at least two epitopes on a protein with at least one antibody being directed toward a post-translationally modified epitope. These antibodies are modified and contacted with a sample as previously described in the present application. The function of proteins are often dependent upon the existence of groups added after primary translation and folding. For example, phosphorylation of tyrosine residues often activate or deactivate a protein kinase or lead to protein-protein dissociation of one of the G-protein subunits in a G-protein coupled receptor. Other protein modification which affect protein function, stability, cellular trafficking or targeting for degradation include, but are not limited to, glycosylation, ADP-ribosylation, myristoylation, palmitoylation and ubiquitination. In the present invention, antibodies can be obtained to epitopes resulting from post-translational modifications of the protein allowing for identification and characterization of post-translational modifications on a protein. Current protein sequencing techniques and oligonucleotides sequencing methodologies do not establish these side chain modifications. In fact, in Edman Degradation, such modifications can make sequencing of a protein very difficult, if not impossible. In addition, in contrast to standard protocols which require a relatively purified sample, a crude cellular homogenate can be used in the present invention since the antibodies are specific to epitopes within a specific protein and the chances of two or more antibodies nonspecifically binding to the same protein are very small.

The following nonlimiting examples are provided for illustrative purpose only.

- 12 -

**EXAMPLES****Example 1: Preparation of antibodies**

Antibodies generated against single, bi- or tri-amino acids are generated by first coupling the antigen to a carrier protein. Examples of carrier proteins include, but are not limited to keyhole limpet hemocyanin and bovine serum albumin. This coupling is accomplished using any number of methodologies well known in the art including, but not limited to, glutaraldehyde, m-Maleimidobenzoyl-N-hydroxysuccinimide ester, carbodiimide or bis-diazotized benzidine. After coupling, the antigen-carrier complex is used as an immunogen to immunize rabbits in accordance with procedures well known in the art. Once appropriate antibody titers are reached, antibodies are isolated from the rabbit blood in accordance with published procedures.

Alternatively, monoclonal antibodies can be generated by immunizing mice with the antigen-carrier complex and generating hybridoma cells lines. For example, the tau monoclonal antibodies used in one embodiment of the invention have been described by Kosik et al. (1988). Ab14 and Abtau1 were derived from a mouse immunization with bovine tau protein. This immunogen was prepared by boiling a thrice-cycled microtubule preparation. The supernatant was fractionated on a Sepharose 4B column to obtain a preparation enriched in tau. In another embodiment, antibodies to phospho-tyrosine are produced. Phosphorylated tyrosine is conjugated to the carrier hemocyanin and antibodies are produced in rabbits.

Animals are injected with the antigen-carrier complex to produce a humoral response and an appropriate screening procedure developed. The sera from test bleed are used to develop and validate the screening procedure. After an appropriate screening procedure has been developed, production of the hybridoma may begin. Several days prior to the fusion, animals are boosted with a sample of antigen. On the day of the fusion, antibody-secreting cells are isolated from lymphoid tissue of the immunized animal and mixed with myeloma cells. The mixture is centrifuged to generate good cell-to-cell

- 13 -

contacts and fused with polyethylene glycol (PEG). Following the fusion, the cells are removed from the PEG, diluted in selective medium and plated into multiwell tissue culture dishes. Samples of the tissue culture supernatants are tested  
5 for the presence of selected antibodies about one week after the fusion. Hybridomas testing positive are grown and then single-cell cloned. Using this approach allows for the establishment of a continuous cell line as the source of a single type of antibody.

10 **Example 2: Preparation of modified Fab fragments**

Fab fragments are prepared from the antibodies in accordance with published procedures using either papain or pepsin digestion of the Ab molecule.

In papain digestion, the antibody is cleaved at the  
15 N-terminal side of the disulfide bond which holds the heavy chains together. If papain digestion is used, a solution of the IgG at 5 mg/ml in 100 mM NaAcetate (pH 5.5) containing cysteine (final concentration 50 mM) and EDTA (final concentration 1 mM) is prepared and aliquoted into test tubes.  
20 Ten micrograms of papain is added for each milligram of antibody in the tube. The tube is then incubated at 37°C for 10 hours. Following incubation, iodoacetamide is added to a final concentration of 75 mM and the solution is incubated for 30 minutes.

25 In pepsin digestion, the antibody is cleaved on the C-terminal side of this bond. If pepsin is used, a solution of IgG at approximately 5 mg/ml in 10 mM sodium citrate, pH 3.5 is prepared and placed in a tube. Five micrograms of pepsin are then added for each milligram of antibody. The tubes are  
30 sealed, mixed well and incubated for 12 to 24 hours at 37°C.

The Fab fragments are separated from the antibodies by sizing over a gel filtration column (G100) and by ion exchange chromatography (DEAE).

- 14 -

**Example 3: Preparation of the DNA-spanner oligonucleotides**

The following oligonucleotides were synthesized by a solid-phase synthesizer.

BAM36-198 SEQ ID NO: 1

5 5' - (CAA)<sub>12</sub>-GGA-TCC- (CAA)<sub>47</sub>-GCT-GAT-GCT-GAT-GCT-3'

BAM36-198op SEQ ID NO: 2

5' - (CAA)<sub>12</sub>-GGA-TCC- (CAA)<sub>47</sub>-GCT-GAT-GCT-GAT-GCT-3'

RI36-198 SEQ ID NO: 3

5' - (CAA)<sub>12</sub>-GAA-TTC- (CAA)<sub>47</sub>-GCT-GAT-GCT-GAT-GCT-3'

10 RI36-198op SEQ ID NO: 4

5' - (CAA)<sub>12</sub>-GAA-TTC- (CAA)<sub>47</sub>-GCT-GAT-GCT-GAT-GCT-3'

Oligonucleotides were phosphorylated in a reaction mixture containing 25 microliters oligo (200 ng/microliter), 5 microliter 10X PNC buffer (500 mM Tris pH 8.0, 10 mM MgCl<sub>2</sub>), 5  
15 microliter ATP (10 mM stock), 10 microliters H<sub>2</sub>O, and 5 microliter polynucleotide kinase. The mixture was incubated for 30 minutes at 37°C. Phenol/chloroform extraction was then performed followed by ethanol precipitation in accordance with well known methods.

20 Derivatization of these oligonucleotides is accomplished through use of the OligoLink derivatization kit provided by Pierce Chemical. The protocol provided by Pierce uses approximately 5 micrograms oligo mixed with a derivatization reagent (cystamine). The derivatized oligo is  
25 then bound to a matrix and reduced with DTT to generate a free disulfide bond in the center of the cystamine leaving a free sulfhydryl group. The oligo in its final form is eluted from the matrix using 10 mM TE (Tris-EDTA) buffer.

**Example 4: Covalent Attachment of DNA-spanner to Antibody**

30 The derivatized oligonucleotides are attached to the antibody using glutaraldehyde. Approximately, 100 micrograms of antibody are mixed with 10 micrograms of the derivatized

- 15 -

oligos in phosphate buffered saline (PBS). A total volume of 200 microliters to 1 ml is appropriate. An equal volume of freshly made 0.1% EM grade glutaraldehyde (in PBS) is added and the mixture incubated at room temperature for 3 hours. 1/20 solution volume of 1 M ethanolamine pH 7 is added to the solution and incubated at room temperature for 2 hours.

The antibody-oligo complexes are adjusted to pH 8.0 by addition of 1 M Tris pH 8 to a final concentration of 0.1 M. The solution is passed over protein A beads (BioRad Corp.) and the column washed with 5 volumes 0.1 M Tris pH 8, then 5 volumes 10 mM Tris pH 8. The antibody-oligo complex is eluted by adding 500 microliters 100 mM glycine pH 3 in 50 microliter aliquots. These aliquots are collected from the column into Eppendorf tubes containing 40 microliters 1 M Tris pH 8. The protein containing fractions are detected by absorbance readings at 280 nm. The protein containing fractions are pooled and stored until use at 4°C in 50% glycerol and 0.02% sodium azide.

**Example 5: Protein antigen interaction with antibody-DNA-spanner complex**

In one embodiment, 10 microliters of 10X binding/reaction buffer (2 M Tris OAc pH 7.4, 1% gelatin, 2% Tween 20, 0.1% sodium azide), 70 microliters H<sub>2</sub>O, 50 ng purified tau protein, 100 ng each Ab14-BAM36-198, Ab14-BAM36-198op, Abtau1-RI36-198, and Abtau1-RI36-198op Ab-DNA-spanner complexes were mixed. Distilled water was added to a final volume of 100 microliters and incubated at room temperature for 30 minutes. 1 microliter 10X binding/reaction buffer, 1.5 microliters 1 M MgOAc, 4 microliters of a mixture of 250 micromolar each dATP, dGTP, dCTP, and TTP, 1 microliter H<sub>2</sub>O, 3 microliters T4 DNA Polymerase (2 U/microliter), 1 microliter Klenow (8 U/microliter) were added and the mixture incubated at room temperature for 30 minutes.

The reactions were then digested with the appropriate restriction enzymes. The ion concentration was adjusted with the addition of 1.5 microliters 1 M KOAc and 3 microliters

- 16 -

restriction enzyme (BamHI and EcoRI). The tubes were then incubated at 37° for 30 minutes.

**Example 6: Analysis of the Digested Complexes**

After the reactions are complete, the samples are  
5 mixed with gel loading dye, heated to 60°C for 5 minutes and  
electrophoresed on a 4% polyacrylamide gel (Laemmli, Nature  
1970, 227: 680-685); 42-1 acrylamide: bis-acrylamide. After  
the bromphenol blue dye reaches the bottom of the gel, the  
complexes are transferred to nitrocellulose membrane by  
10 electroblotting.

The nitrocellulose is baked at 80°C under vacuum for  
4 hours. The blot is prehybridized for 4 hours at 40°C in 6X  
SSC (0.9 M sodium citrate, 0.09 M sodium chloride); 100  
microgram/ml salmon sperm DNA, 5X Denhardt's (0.1% BSA, 0.1%  
15 polyvinylpyrrolidone, 0.1% ficoll) and 50% formamide. The blot  
is then hybridized at 50°C for 18 hours with <sup>32</sup>P-radiolabeled  
(TGG)<sub>10</sub>. This probe is labeled using 5'-polynucleotide kinase  
where the reaction volume is 10 microliters and 3 microliters  
of <sup>32</sup>P-ATP (3000 Ci/mM, Amersham) is utilized in the reaction  
20 mix. After hybridization, the blot is washed twice for 30  
minutes each with 2X SSC at 50°C followed by washing twice with  
0.1X SSC at 50°C for 30 minutes each. The blot is exposed to  
X-ray film next to an intensifying screen for 24 hours.

- 17 -

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: James Eberwine
- (ii) TITLE OF INVENTION: A Method of Sequencing Proteins by  
Epitope Ordering and Protein  
Restriction Mapping
- (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Jane Massey Licata, Esq.
  - (B) STREET: 210 Lake Drive East, Suite 201
  - (C) CITY: Cherry Hill
  - (D) STATE: NJ
  - (E) COUNTRY: USA
  - (F) ZIP: 08002
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
  - (B) COMPUTER: IBM 486
  - (C) OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
  - (D) SOFTWARE: WORDPERFECT 5.1
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: not yet assigned
  - (B) FILING DATE: Herewith
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: 08/294,133
  - (B) FILING DATE: August 22, 1994
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Jane Massey Licata
  - (B) REGISTRATION NUMBER: 32,257

- 18 -

(C) REFERENCE/DOCKET NUMBER: PENN-0137

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (609) 779-2400

(B) TELEFAX: (609) 779-8488

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 198

(B) TYPE: Nucleic Acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(iv) ANTI-SENSE: No

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CAACAACAAC AACACAACA ACAACAACAA CAACAAGGAT CCCAACAACA 50  
ACAACAACAA CAACAACAAC AACACAACA ACAACAACAA CAACAACAAC 100  
AACACAACA ACAACAACAA CAACAACAAC AACACAACA ACAACAACAA 150  
CAACAACAAC AACACAACA ACAACAACAA CAAGCTGATG CTGATGCT 198

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 198

(B) TYPE: Nucleic Acid

(C) STRANDEDNESS: Single

(D) TOPOLOGY: Linear

(iv) ANTI-SENSE: No

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CAACAACAAC AACACAACA ACAACAACAA CAACAAGGAT CCCAACAACA 50  
ACAACAACAA CAACAACAAC AACACAACA ACAACAACAA CAACAACAAC 100  
AACACAACA ACAACAACAA CAACAACAAC AACACAACA ACAACAACAA 150  
CAACAACAAC AACACAACA ACAACAACAA CAAGCTGATG CTGATGCT 198

- 19 -

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 198
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(iv) ANTI-SENSE: No

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

CAACAACAAC AACACAACA ACAACAACAA CAACAAGAAT TCCAACAACA 50  
ACAACAACAA CAACAACAAC AACACAACA ACAACAACAA CAACAACAAC 100  
AACACAACA ACAACAACAA CAACAACAAC AACACAACA ACAACAACAA 150  
CAACAACAAC AACACAACA ACAACAACAA CAAGCTGATG CTGATGCT 198

## (2) INFORMATION FOR SEQ ID NO: 4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 198
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(iv) ANTI-SENSE: No

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CAACAACAAC AACACAACA ACAACAACAA CAACAAGAAT TCCAACAACA 50  
ACAACAACAA CAACAACAAC AACACAACA ACAACAACAA CAACAACAAC 100  
AACACAACA ACAACAACAA CAACAACAAC AACACAACA ACAACAACAA 150  
CAACAACAAC AACACAACA ACAACAACAA CAAGCTGATG CTGATGCT 198

- 20 -

**What is claimed is:**

1. A method of identifying an amino acid sequence of a protein comprising:

- a) obtaining antibodies to amino acids;
- 5 b) modifying said antibodies by binding a first nucleotide sequence and a second nucleotide sequence to said antibody, said first nucleotide sequence comprising a modified 5'-nucleotide followed by a single restriction enzyme site and a long nucleotide sequence containing a base pair sequence at  
10 the 3' end capable of annealing to a complementary sequence, said second nucleotide sequence comprising the same modified 5'-nucleotide, single restriction enzyme site and a long nucleotide sequence as said first nucleotide sequence and having the complementary base pair sequence at the 3' end;
- 15 c) binding modified antibodies to a protein so that complementary base pair sequences at the 3' end of said first and second nucleotide sequences of said modified antibodies form a primer template complex which is extended to form a stable double-stranded duplex;
- 20 d) treating the double-stranded duplex comprising bound modified antibodies with at least one restriction enzyme to digest the duplex;
- e) separating products of the digestion;
- f) detecting a band pattern produced from these  
25 digested products; and
- g) identifying a portion of the amino acid sequence of the protein.

2. The method of claim 1 wherein said antibodies are Fab fragments.

30 3. The method of claim 1 wherein the long nucleotide sequence of the first and second nucleotide sequences comprises 200 base pairs.

- 21 -

4. The method of claim 1 wherein the base pair sequences at the 3' end of said first and second nucleotide sequences comprises 14 base pairs.

5. A method of characterizing post-translational events on a protein comprising:

- a) obtaining antibodies to at least two epitopes on a protein wherein at least one antibody is directed a post-translationally modified epitope;
- b) modifying said antibodies by binding a first  
10 nucleotide sequence and a second nucleotide sequence to said antibody, said first nucleotide sequence comprising a modified 5'-nucleotide followed by a single restriction enzyme site and a long nucleotide sequence containing a base pair sequence at the 3' end capable of annealing to a complementary sequence,  
15 said second nucleotide sequence comprising the same modified 5'-nucleotide, single restriction enzyme site and a long nucleotide sequence as said first nucleotide sequence and having the complementary base pair sequence at the 3' end;
- c) binding modified antibodies to a protein so that  
20 complementary base pair sequences at the 3' end of said first and second nucleotide sequences of said modified antibodies form a primer template complex which is extended to form a stable double-stranded duplex;
- d) treating the double-stranded duplex comprising bound  
25 modified antibodies with at least one restriction enzyme to digest the duplex;
- e) separating products of the digestion;
- f) detecting a band pattern produced from these digested products; and
- 30 g) identifying and characterizing a post-translation event on the protein.

6. The method of claim 5 wherein said antibodies are Fab fragments.

- 22 -

7. The method of claim 5 wherein the long nucleotide sequence of the first and second nucleotide sequences comprises 200 base pairs.

8. The method of claim 5 wherein the base pair sequences 5 at the 3' end of said first and second nucleotide sequences comprises 14 base pairs.

9. A modified antibody comprising:

a) an antibody to an amino acid;

b) a first nucleotide sequence having a modified 5'-  
10 nucleotide, a single restriction enzyme site, a long nucleotide sequence, and a first base pair sequence at the 3' end which is capable of annealing to a complementary sequence, said first nucleotide sequence being bound to said antibody; and

15 c) a second nucleotide sequence having said modified 5'-nucleotide, said single restriction enzyme site and said long nucleotide sequence as in said first nucleotide sequence and a second base pair sequence at the 3' end which is complementary to said first base pair sequence of said first  
20 nucleotide sequence, said second nucleotide sequence being bound to said antibody.

10. The modified antibody of claim 9 wherein the antibody comprises a Fab fragment.

11. The modified antibody of claim 9 wherein said long  
25 nucleotide sequence comprises 200 base pairs.

12. The modified antibody of claim 9 where said first and second base pair sequences each comprise 14 base pairs.

13. A modified antibody comprising:

30 a) an antibody to an post-translationally modified epitope of a protein;

- 23 -

b) a first nucleotide sequence having a modified 5'-nucleotide, a single restriction enzyme site, a long nucleotide sequence, and a first base pair sequence at the 3' end which is capable of annealing to a complementary sequence,  
5 said first nucleotide sequence being bound to said antibody;  
and

c) a second nucleotide sequence having said modified 5'-nucleotide, said single restriction enzyme site and said long nucleotide sequence as in said first nucleotide sequence  
10 and a second base pair sequence at the 3' end which is complementary to said first base pair sequence of said first nucleotide sequence, said second nucleotide sequence being bound to said antibody.

14. The modified antibody of claim 13 wherein the antibody  
15 comprises a Fab fragment.

15. The modified antibody of claim 13 wherein said long nucleotide sequence comprises 200 base pairs.

16. The modified antibody of claim 13 where said first and second base pair sequence each comprise 14 base pairs.

20

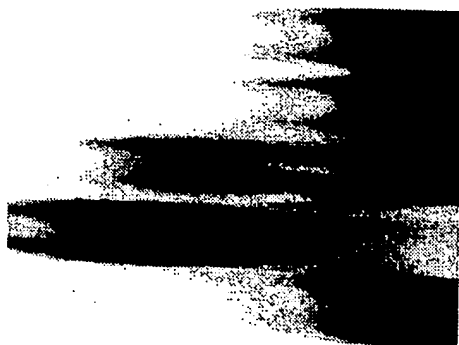


FIG. 1A

Ab14-DNA-SPANNER —  
— SPANNER DNA-Abtau1

Ab14-DNA-SPANNER  
+  
Abtau1-DNA-SPANNER



LANE 2



LANE 3

Ab14-DNA-SPANNER BamHI CUT

EcoRI CUT-SPANNER DNA-Abtau1



LANE 4



LANE 5

FIG. 1B

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US95/10668

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : A61K 35/16, 39/00, 39/395

US CL : Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/7.1, 91.1, 183; 530/387.1, 391.3, 866; 536/24.33; 935/59, 76, 77

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MEDLINE, WPI AND APS

SEARCH TERMS: AMINO ACID, ANTIBODY, NUCLEOTIDE, PRIMER, SEQUENCE, PROTEIN, PEPTIDE, AND INVENTOR NAMES

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
L	REES ET AL. "Protein Engineering: A Practical Approach" Published 1992 by IRL Press (NY), Page 1, Chapter 11, (This reference discusses the limitations of the antibody binding region).	1-16
L	METZLER "BIOCHEMISTRY: The Chemical Reactions of Living Cells", Published 1977 by Academic Press, Inc. (NY), see page 64, (this reference teaches the size of a tri-peptide).	1-16

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	*Y	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z	document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means		
*P* document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

30 OCTOBER 1995

Date of mailing of the international search report

27 NOV 1995

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

DONALD E. ADAMS, PH.D.

Telephone No. (703) 308-0196

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US95/10668

## A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

435/7.1, 91.1, 183; 530/387.1, 391.3, 866; 536/24.33; 935/59, 76, 77